Bayesian approach to cubic natural exponential families

Marwa Hamza and Abdelhamid Hassairi *

Laboratory of Probability and Statistics. Sfax Faculty of Sciences, B.P. 1171, Tunisia.

Abstract For a natural exponential family (NEF), one can associate in a natural way two standard families of conjugate priors, one on the natural parameter and the other on the mean parameter. These families of conjugate priors have been used to establish some remarkable properties and characterization results of the quadratic NEF's. In the present paper, we show that for a NEF, we can associate a class of NEF's, and for each one of these NEF's, we define a family of conjugate priors on the natural parameter and a family of conjugate priors on the mean parameter which are different of the standard ones. These families are then used to extend to the Letac-Mora class of real cubic natural exponential families the properties and characterization results related to the Bayesian theory established for the quadratic natural exponential families.

Keywords: natural exponential family, variance function, prior distribution, posterior expectation.

1 Introduction and preliminaries

To make clear the motivations of the present paper, we first recall some facts concerning the natural exponential families and their variance functions. Our notations are the ones used by Letac in [9]. Let μ be a positive radon measure on $\mathbb{R}$, and denote by

$$L_\mu(\lambda) = \int \exp(\lambda x)\mu(dx) \tag{1.1}$$

its Laplace transform. Let $\mathcal{M}(\mathbb{R})$ be the set of measures μ such that the set

$$\Theta(\mu) = \text{interior}\{\lambda \in \mathbb{R};\ L_\mu(\lambda) < +\infty\} \tag{1.2}$$

is non empty and μ is not Dirac measure. The cumulant function of an element μ of $\mathcal{M}(\mathbb{R})$ is the function defined for $\lambda \in \Theta(\mu)$ by

$$k_\mu(\lambda) = \ln L_\mu(\lambda).$$

*Corresponding author. *E-mail address: Abdelhamid.Hassairi@fss.rnu.tn*

To each μ in $\mathcal{M}(\mathbb{R})$ and λ in $\Theta(\mu)$, we associate the probability distribution on $\mathbb{R}$

$$P(\lambda,\mu)(dx) = \exp(\lambda x - k_\mu(\lambda))\mu(dx). \tag{1.3}$$

The set

$$F(\mu) = \{P(\lambda,\mu);\ \lambda \in \Theta(\mu)\}$$

is called the natural exponential family (NEF) generated by μ.

The function k_μ is strictly convex and analytic. Its first derivative k'_μ defines a diffeomorphism between $\Theta(\mu)$ and its image $M_{F(\mu)}$. Since $k'_\mu(\lambda) = \int x\ P(\lambda,\mu)(dx)$, $M_{F(\mu)}$ is called the domain of the means of $F(\mu)$. The inverse function of k'_μ is denoted by ψ_μ and setting

$$P(m, F(\mu)) = P(\psi_\mu(m), \mu) \tag{1.4}$$

the probability of $F(\mu)$ with mean m, we have

$$F(\mu) = \left\{P(m, F(\mu));\ m \in M_{F(\mu)}\right\},$$

which is the parametrization of $F(\mu)$ by the mean.
The variance of $P(m, F(\mu))$ is denoted by $V_{F(\mu)}(m)$ and the map defined from $M_{F(\mu)}$ into $L_s(\mathbb{R})$, the set of symmetric function in $\mathbb{R}$, by

$$m \longmapsto V_{F(\mu)}(m) = k''_\mu(\psi_\mu(m)) = (\psi'_\mu(m))^{-1}$$

is called the variance function of the NEF $F(\mu)$ generated by μ. We also say that μ is a basis of $F(\mu)$. An important feature of $V_{F(\mu)}$ is that it characterizes the natural exponential family $F(\mu)$ in the following sense: If $F(\mu)$ and $F(\nu)$ are two NEF's such that $V_{F(\mu)}(m)$ and $V_{F(\nu)}(m)$ coincide on a nonempty open subset of $M_{F(\mu)} \cap M_{F(\nu)}$, then $F(\mu) = F(\nu)$.
Now, for $\mu \in \mathcal{M}(\mathbb{R})$ the Jørgensen set of μ or of $F(\mu)$ is defined by

$$\Lambda(\mu) = \{\lambda > 0;\ \exists\ \mu_\lambda;\ L_{\mu_\lambda}(\theta) = (L_\mu(\theta))^\lambda \ \textit{and}\ \Theta(\mu_\lambda) = \Theta(\mu)\}.$$

$\Lambda(\mu)$ is stable under addition which means that if λ and λ' are in $\Lambda(\mu)$, then $\lambda + \lambda'$ are in $\Lambda(\mu)$, and $\mu_{\lambda+\lambda'} = \mu_\lambda * \mu_{\lambda'}$.
For all λ in $\Lambda(\mu)$ we have

$$M_{F(\mu_\lambda)} = \lambda M_{F(\mu)} \ \ \textit{and}\ \ V_{F(\mu_\lambda)}(m) = \lambda V_{F(\mu)}(\frac{m}{\lambda}).$$

Several classifications of NEFs based on the form of the variance function have been realized in the last three decades. The most interesting classes of real NEF's are the class of quadratic NEFs, i.e., the class of NEF's such that the variance function is a polynomial in the mean of degree less than or equal to 2 characterized by Morris [11], and the class of cubic NEF's, i.e., the class of NEF's such that the variance function is a polynomial in the mean of degree less than or equal to 3 characterized by Letac and Mora [10]. Recall that up to affine transformations and power of convolution the class of quadratic NEF's contains six families: the gaussian, the Poisson, the gamma, the binomial, the negative binomial, an the hyperbolic family. The class of cubic NEF's

contains, besides the quadratic ones, six other families, the most famous is the inverse Gaussian distribution with variance function $V(m) = m^3$. It is worth mentioning here that multivariate versions of these classes have also been defined and completely described. For instance, Casalis [1] has described the so-called class of multivariate simple quadratic NEFs and Hassairi [6] has described the class of multivariate simple cubic NEFs which are respectively the generalizations of the real quadratic and real cubic NEF's. The fact that the variance function of a family is quadratic or cubic is not only a question of form but it corresponds to some interesting analytical characteristic properties. Indeed, the Morris class of quadratic NEF's has some characterizations involving orthogonal polynomials due to Fiensilver [5]. These characterizations have been extended to the Letac and Mora class of real cubic NEF's by Hassairi and Zarai [8] using a notion of 2-orthogonality of a sequence of polynomials. Other remarkable characterizations of the quadratic NEF's are related to the Bayesian theory. For instance, given a NEF $F(\mu)$, Diaconis and Ylvisaker [4] have considered the standard family Π of priors on the natural parameter λ defined by

$$\pi_{t_1,m_1}(d\lambda) = C_{t_1,m_1}\ \exp(t_1 m_1 \lambda - t_1 k_\mu(\lambda))\ \mathbf{1}_{\boldsymbol{\Theta}(\mu)}(\lambda)\ d\lambda \tag{1.5}$$

where $t_1 > 0$, m_1 is in $M_{F(\mu)}$, and C_{t_1,m_1} is a normalizing constant. This distribution is in fact a particular case of the so called implicit distribution on the parameter of a statistical model introduced in [7]. They have shown that if X is a random variable distributed according to $P(\lambda, \mu)$ (see (1.3)), then the only conjugate family of prior distributions on λ that gives a linear posterior expectation of $k'_\mu(\lambda)$ given X is the standard one Π (see also [2]). Consonni and Veronese [3] have considered another family $\widetilde{\Pi}$ of prior distributions $\widetilde{\pi}_{t_1,m_1}$ on the mean parameter m defined also for $t_1 > 0$ and m_1 in $M_{F(\mu)}$ by

$$\widetilde{\pi}_{t_1,m_1}(dm) = \widetilde{C}_{t_1,m_1} \exp(t_1 m_1 \psi_\mu(m) - t_1 k_\mu(\psi_\mu(m)))\ \mathbf{1}_{M_{F(\mu)}}(m)\, dm. \tag{1.6}$$

They have shown that the fact that $\widetilde{\Pi}$ contains $k'_\mu(\Pi)$ characterizes the quadratic NEFs. These authors have also shown that if the prior on the mean parameter m is $\widetilde{\pi}_{t_1,m_1}$, then under some conditions on the support of μ, the NEF $F(\mu)$ is quadratic if and only if the posterior expectation of $k'_\mu(\lambda)$ is a linear function of the sample mean. We also mention that Diaconis and Yilvisaker [4] have shown that if the standard prior on λ is given by π_{t_1,m_1} with $t_1 > 0$ and m_1 is in $M_{F(\mu)}$, then the expectation of $k'_\mu(\lambda)$ is equal to m_1, that is

$$C_{t_1,m_1} \int k'_\mu(\lambda)\ \exp(t_1 m_1 \lambda - t_1 k_\mu(\lambda))\ \mathbf{1}_{\boldsymbol{\Theta}(\mu)}(\lambda) d\lambda = m_{\mathbf{1}}, \tag{1.7}$$

or equivalently in terms of the mean parameter

$$C_{t_1,m_1} \int \frac{m}{V_{F(\mu)}(m)} \exp(t_1 m_1 \psi_\mu(m) - t_1 k_\mu(\psi_\mu(m)))\ \mathbf{1}_{M_{F(\mu)}}(m) dm = m_{\mathbf{1}}. \tag{1.8}$$

A natural question within this approach is if one can extend the properties and characterization results concerning the quadratic NEF's and related to the Bayesian theory to the Letac-Mora class of real cubic NEFs. The aim of the present paper is to give an answer to this question. We first introduce, for a given NEF $F(\nu)$ and β in some interval of $\mathbb{R}$ containing 0, a NEF $F^\beta(\nu)$ such that $F^0(\nu) = F(\nu)$. We then define a family Π^β of prior distributions on the natural parameter θ and a family $\widetilde{\Pi}^\beta$ of prior distributions on the mean parameter m which may be seen as generalizations of the families Π and $\widetilde{\Pi}$ defined

above, since $\Pi = \Pi^0$ and $\widetilde{\Pi} = \widetilde{\Pi}^0$. After proving that for each β, the family $\widetilde{\Pi}^\beta$ is a conjugate family of prior distributions with respect to the NEF $F^\beta(\nu)$, we show that a cubic NEF $F(\nu)$ is characterized by the fact that there exists a β such that the posterior expectation of $\dfrac{k'_\nu(\theta)}{1-\beta k'_\nu(\theta)}$ is linear when the prior on θ is π^β_{t,m_0}. We also show that a cubic NEF $F(\nu)$ is characterized by a differential equation verified by the cumulant function k_ν. A third characterization of a real cubic NEF is realized when the family of priors $\widetilde{\Pi}^\beta$ contains the family $k'_\nu(\Pi^\beta)$. The restriction of all these results to the subclass of quadratic NEF's leads to the results of Diconis and Ylvisaker[4] and Consonni and Veronese[3]. The results of the paper are illustrated by an example.

2 Main results

In this section, we state and prove our main results. Our considerations will be restricted to regular NEFs, so that the domain of the means of a NEF is equal to the interior of the convex hull of its support. This property of regularity is satisfied by all the most common NEF's. An important fact which will be crucial in our proofs is that up to affine transformations and powers of convolution, a cubic natural exponential family may be obtained from a quadratic one by the so-called action of the linear group $GL(\mathbb{R}^2)$ on the real families. Originally, the action of the linear group includes the affine transformations and powers of convolution, however since these transformations preserve the class of quadratic NEF's and the class of cubic NEF's, we will focus on the facts which we need here, for more precise statements in this connection, refer to Hassairi[6]. Let $F(\nu)$ and $F(\mu)$ be two real NEFS's. Suppose that there exists a β in $\mathbb{R}$ such that the set

$$(M_{F(\nu)})_\beta = \{m \in M_{F(\nu)}; 1+\beta m > 0\}$$

is nonempty and for m in $(M_F)_\beta$,

$$V_{F(\nu)}(m) = (1+\beta m)^3 \; V_{F(\mu)}\left(\frac{m}{1+\beta m}\right) \;, \tag{2.9}$$

then we write $F(\nu) = T_\beta(F(\mu))$. This defines an action on the natural exponential families, so that we have $T_\beta T_{\beta'} = T_{\beta+\beta'}$ and $F(\nu) = T_\beta(F(\mu))$ is equivalent to $F(\mu) = T_{-\beta}(F(\nu))$. We also mention that $F(\nu) = T_\beta(F(\mu))$ may be expressed in terms of the cumulant functions of the generating measures by

$$\left\{\begin{array}{lcl} k_\nu(\theta) & = & k_\mu(\lambda) \\ \\ \theta & = & -\beta k_\mu(\lambda)+\lambda \end{array}\right. \tag{2.10}$$

or equivalently by

$$\left\{\begin{array}{lcl} k_\mu(\lambda) & = & k_\nu(\theta) \\ \\ \lambda & = & \beta k_\nu(\theta)+\theta \end{array}\right. \tag{2.11}$$

An important fact is that the relation (2.11) between the cumulent functions may be explicitly given it terms of the measures themselves. In fact if the α-power of convolution

ν_α of ν is written as $\nu_\alpha(dx) = h(\alpha, x)\sigma(dx)$, where $\sigma(dx)$ is either Lebesgue measure or a counting measure, then the measure

$$\mu(dx) = \frac{1}{1-\beta x} h(1-\beta x, x)\mathbf{1}_{\mathbf{\Lambda}(\nu)}(1-\beta x)\sigma(dx)$$

satisfies (2.10) and generates the family $T_{-\beta}(F(\nu))$. This measure μ will be denoted $T_{-\beta}(\nu)$ and the family $F(\mu) = T_{-\beta}(F(\nu))$ will be denoted F^β.
We mention here that if $F(\nu)$ is a cubic NEF, there exists β in $B_{F(\nu)}$ and a quadratic NEF $F(\mu)$ such that $F(\nu) = F(T_\beta(\mu))$ or equivalently, $F(\mu) = F(T_{-\beta}(\nu))$.
Besides the restriction to half lines for the domain of the means in the definition of $(M_{F(\nu)})_\beta$, we also define for $\nu \in \mathcal{M}(\mathbb{R})$ and $\beta \in \mathbb{R}$, the sets

$$H_\beta(\nu) = \{x \in \mathbb{R};\ 1 + \beta x \in \Lambda(\nu)\},$$

and

$$B_{F(\nu)} = \{\beta \in \mathbb{R};\ \nu(H_\beta) > 0\}.$$

We have the following preliminary result.

Proposition 2.1 *Let $F(\nu)$ be a regular NEF. Then*

$$\beta \in B_{F(\nu)} \quad if \quad and \quad only \quad if \quad (M_{F(\nu)})_\beta \neq \emptyset.$$

Proof We will make a reasoning for $\beta \geq 0$, the case $\beta < 0$ may be done in a similar way. Suppose that there exists m_0 in $(M_{F(\nu)})_\beta$, that is $m_0 \in M_{F(\nu)}$ and $1 + \beta m_0 > 0$. As $M_{F(\nu)}$ is equal to the interior of the convex hull of *supp*(ν), there exist x_0 in *supp*(ν) such that $x_0 > m_0$. This with the fact that $1 + \beta m_0 > 0$ imply that $1 + \beta x_0 > 0$. Thus H_β is an open set which contains an element of *supp*(μ). It follows that $\nu(H_\beta) > 0$ and β is in $B_{F(\nu)}$.
Conversely, if β is in $B_{F(\nu)}$, then $\nu(H_\beta) > 0$. Since H_β is an open set, this implies that it contains an element x_0 of *supp*(ν). We have on the one hand that $1 + \beta x_0 > 0$ so that there exists $\varepsilon > 0$ such that $1 + \beta x_0 - \beta\varepsilon > 0$. On the other hand, as $M_{F(\nu)}$ is equal to the interior of the convex hull of *supp*(ν), there exists m_0 in $(M_{F(\nu)})$ such that $|m_0 - x_0| < \varepsilon$. From this we deduce that m_0 is in $(M_{F(\nu)})_\beta$. □

We now use the natural parametrization and the parametrization by the mean of the original family $F(\nu)$ to give two parameterizations of the family $F(\mu)$. These parameterizations are, for $\beta \neq 0$, different of the usual parameterizations of $F(\mu)$. In fact, for $\theta \in \Theta(\nu)$, we write

$$P(\beta, \theta, \nu)(dx) = \exp\{(\theta + \beta k_\nu(\theta))x - k_\nu(\theta)\}\ T_{-\beta}(\nu)(dx).$$

Similarly, parameterizing by $m \in M_{F(\nu)}$, we write

$$P(\beta, m, F(\nu))(dx) = \exp\{(\psi_\nu(m) + \beta k_\nu(\psi_\nu(m)))x - k_\nu(\psi_\nu(m))\}\ T_{-\beta}(\nu)(dx).$$

Thus we have that

$$F(\mu) = F^\beta = \{P(\beta, \theta, \nu)(dx);\ \theta \in \Theta(\nu)\} = \{P(\beta, m, F(\nu))(dx);\ m \in M_{F(\nu)}\}.$$

Accordingly, we define for β in $B_{F(\nu)}$ two families of prior distributions. Let

$$(\Theta)_\beta = \{\theta \in \Theta(\nu);\ 1 + \beta k'_\nu(\theta) > 0\}.$$

Then we have that $(M_{F(\nu)})_\beta = k'_\nu((\Theta)_\beta)$, and we define for $t>0$ *and* $m_0 \in (M_{F(\nu)})_\beta$,

$$\pi^\beta_{t,m_0}(d\theta) = C^\beta_{t,m_0}\ (1+\beta k'_\nu(\theta))\ \exp(tm_0\theta - tk_\nu(\theta))\ \mathbf{1}_{(\boldsymbol{\Theta})_\beta}(\theta)d\theta, \tag{2.12}$$

and

$$\Pi^\beta = \{\pi^\beta_{t,m_0};\ t>0 \ \textit{and}\ m_0 \in (M_{F(\nu)})_\beta\},$$

This family comes in fact from the standard family Π defined in (1.5) using (2.11). The normalizing constant C^β_{t,m_0} is then well defined for $t>0$ *and* $m_0 \in (M_{F(\nu)})_\beta$.
Besides this family of priors on the parameter θ, we define a family of priors on the parameter m. Always for $t>0$ *and* $m_0 \in (M_{F(\nu)})_\beta$, we consider the probability distribution

$$\widetilde{\pi}^\beta_{t,m_0}(dm) = \widetilde{C}^\beta_{t,m_0}\ (1+\beta m)^{-2}\ \exp(tm_0\psi_\nu(m) - tk_\nu(\psi_\nu(m)))\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)dm,$$

where $\widetilde{C}^\beta_{t,m_0}$ is a normalizing constant. It is the image of $\widetilde{\pi}_{t_1,m_1}$ defined in (1.6) by the map $m' \longmapsto \dfrac{m'}{1-\beta m'}$. The family of priors on m is then

$$\widetilde{\Pi}^\beta = \{\widetilde{\pi}^\beta_{t,m_0};\ t>0 \ \textit{and}\ m_0 \in (M_{F(\nu)})_\beta\},$$

Next, we prove that these families are conjugate families of prior distributions.

Proposition 2.2 *i) The family Π^β is conjugate family of prior distributions with respect to the NEF F^β parameterized by the natural parameter θ.*
ii) The family $\widetilde{\Pi}^\beta$ is a conjugate family of prior distributions with respect to the NEF F^β parameterized by the mean parameter m.

Proof i) Suppose that X is a random variable with distribution $P(\beta,\theta,\nu)(dx)$ and that π^β_{t,m_0} is a prior distribution on the parameter θ. Then the posterior distribution is

$$\frac{C^\beta_{t,m_0}\ (1+\beta k'_\nu(\theta))\ \exp((tm_0+x)\theta - (t+1-\beta x)k_\mu(\theta))\ \mathbf{1}_{(\boldsymbol{\Theta})_\beta}(\theta)}{\displaystyle\int C^\beta_{t,m_0}\ (1+\beta k'_\nu(\theta))\ \exp((tm_0+x)\theta - (t+1-\beta x)k_\mu(\theta))\ \mathbf{1}_{(\boldsymbol{\Theta})_\beta}(\theta)\ d\theta}.$$

If we set $t_2 = t+1-\beta x$ and $m_2 = \dfrac{tm_0+x}{t+1-\beta x}$, then this distribution is noting but $\pi^\beta_{t+1-\beta X,(tm_0+X)/(t+1-\beta X)}$, and it belongs to Π^β. In fact, since $t>0$, $T_{-\beta}(\nu)$ is concentrated on $\{1-\beta x>0\}$, and m_0 is in $(M_{F(\nu)})_\beta$, we have that $t_2>0$ and $1+\beta m_2 = \dfrac{1+t(1+\beta m_0)}{t+1-\beta X} > 0$, so that m_2 is in $(M_F)_\beta$.
ii) Suppose now that Y is a random variable $P(\beta,m,F(\nu))$ distributed and that the prior on the mean parameter m is $\widetilde{\pi}^\beta_{t,m_0}$, then the posterior distribution of m is given by

$$\frac{\widetilde{C}^\beta_{t,m_0}\ (1+\beta m)^{-2}\ \exp((tm_0+y)\psi_\mu(m) - (t+1-\beta y)k_\mu(\psi_\mu(m)))\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)}{\displaystyle\int \widetilde{C}^\beta_{t,m_0}\ (1+\beta m)^{-2}\ \exp((tm_0+y)\psi_\mu(m) - (t+1-\beta y)k_\mu(\psi_\mu(m)))\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)\ dm}.$$

This with $t_3 = t+1-\beta y$ and $m_3 = \dfrac{tm_0+y}{t+1-\beta y}$, is equal to $\widetilde{\pi}^\beta_{t+1-\beta Y,(tm_0+Y)/(t+1-\beta Y)}$, with the required conditions to belong to $\widetilde{\Pi}^\beta$. $\square$

Corollary 2.3 *Let $(X_1, ..., X_n)$ be a sample $P(\beta, \theta, \nu)$-distributed and consider π^β_{t,m_0} as the prior on θ. Then the posterior distribution of θ given $X_1, ..., X_n$ is*

$$\pi^\beta_{t+n-\beta n\overline{X},(tm_0+n\overline{X})/(t+n-\beta n\overline{X})}.$$

Proof It is easy to see that the distribution of the random vector $(\theta, X_1, ..., X_n)$ is

$$C^\beta_{t,m_0}\ (1+\beta k'_\nu(\theta))\ \exp((tm_0+n\overline{x})\theta-(t+n(1-\beta\overline{x}))k_\nu(\theta))\ \mathbf{1}_{(\mathbf{\Theta})_\beta}(\theta)\ \prod_{i=1}^{n} T_{-\beta}(\nu)(dx_i)d\theta.$$

With the same technic used in Proposition 2.2, we deduce that the posterior distribution of θ given $X_1, ..., X_n$ is

$$\pi^\beta_{t+n-\beta n\overline{X},(tm_0+n\overline{X})/(t+n-\beta n\overline{X})}.$$

□

Proposition 2.4 *Let $F(\nu)$ be a cubic NEF. Then there exists β in $B_{F(\nu)}$ such that, if the prior on θ is π^β_{t,m_0}, then*

$$E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)}\right)=\frac{m_0}{1+\beta m_0}.$$

Proof Since $F(\nu)$ is cubic then there exist β in $B_{F(\nu)}$ and a quadratic NEF $F(\mu)$ such that $F(\nu)=T_\beta(F(\mu))$ and $\nu=T_\beta(\mu)$. Using (2.11) it is easy to see that if the prior on θ in $(\Theta)_\beta$ is π^β_{t,m_0} then the prior of λ in $\Theta(\mu)$ is the standard π_{t_1,m_1} with $t_1=t(1+\beta m_0)$ and $m_1=\dfrac{m_0}{1+\beta m_0}$. Moreover we have $C^\beta_{t,m_0}=C_{t(1+\beta m_0),m_0/(1+\beta m_0)}$.
It follow that

$$\begin{aligned}
E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)}\right) &= \int C^\beta_{t,m_0}\ \frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)}\ (1+\beta k'_\nu(\theta))\ \exp(tm_0\theta-tk_\nu(\theta))\ \mathbf{1}_{(\mathbf{\Theta})_\beta}(\theta)d\theta\\
&= \int C^\beta_{t,m_0}\ k'_\mu(\lambda)\ \exp(tm_0(\lambda-\beta k_\mu(\lambda))-tk_\mu(\lambda))\ \mathbf{1}_{\mathbf{\Theta}(\mu)}(\lambda)d\lambda\\
&= \int\ C_{t_1,m_1}k'_\mu(\lambda)\ \exp(t_1m_1\lambda-t_1k_\mu(\lambda))\ \mathbf{1}_{\mathbf{\Theta}(\mu)}(\lambda)d\lambda.
\end{aligned}$$

Invoking (1.7) we get

$$E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)}\right)=\frac{m_0}{1+\beta m_0}.$$

□

Now we give a characterization of the cubic NEFs which is based on the linearity of the posterior expectation.

Theorem 2.5 *Let ν be in $M(I\!R)$.*

1. *If $F(\nu)$ is cubic then there exists β in $B_{F(\nu)}$ such that for all $n\geq 1$, if $X_1, ..., X_n$ is a sample with distribution $P(\beta, \theta, \nu)$ and the prior on the natural parameter θ is π^β_{t,m_0}, then $E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)}\mid X_1, ..., X_n\right)$ is linear.*

2. *The converse is true if we assume that $supp(T_{-\beta}(\nu))$ contains an open interval in $I\!R$ or is a finite or denumerably infinite subset of $]-\infty, 0[$, alternatively $]0, +\infty[$, with $T_{-\beta}(\nu)\{0\} > 0$.*

Proof

1. From Proposition 2.4 and Corollary 2.3, we easily get

$$E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)} \mid X_1, ..., X_n\right) = \frac{tm_0 + n\overline{X}}{t(1+\beta m_0)+n},$$

which is linear in $\overline{X}$.

2. Conversely, suppose that there exists β in $B_{F(\nu)}$ such that $E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)} \mid X_1, ..., X_n\right)$ is linear for all n and a sample $X_1, ..., X_n$ with distribution $P(\beta, \theta, \nu)(dx)$. Consider $\mu = T_{-\beta}(\nu)$. Then using (2.11), we have that the distribution $P(\beta, \theta, \nu)(dx)$ of $X_1, ..., X_n$ is equal to $P(\lambda, \mu)(dx)$ which is nothing but an other parametrization involving μ. We also have that

$$E\left(k'_\mu(\lambda) \mid X_1, ..., X_n\right) = E\left(\frac{k'_\nu(\theta)}{1+\beta k'_\nu(\theta)} \mid X_1, ..., X_n\right)$$

which is linear in $X_1, ..., X_n$, from the hypothesis.
On the other hand, as the prior on the natural parameter θ is assumed to be π^β_{t,m_0}, we get as prior on $\lambda \in \Theta(\mu)$ the standard prior distribution given by

$$\pi_{t_1,m_1}(d\lambda) = C_{t_1,m_1}\ \exp(t_1 m_1 \lambda - t_1 k_\mu(\lambda))\ \mathbf{1}_{\Theta(\mu)}(\lambda)(d\lambda),$$

with $t_1 = t(1+\beta m_0) > 0$, $m_1 = \dfrac{m_0}{1+\beta m_0} \in M_{F(\mu)}$ and $C_{t_1,m_1} = C^\beta_{t,m_0}$.
As we have that $supp(\mu) = supp(T_{-\beta}(\nu)) \subset supp(\nu) \cap \{1 - \beta x \in \Lambda(\nu)\}$. the assumptions on $supp(T_{-\beta}(\nu))$ imply that $supp(\mu)$ satisfies hypotheses (H1) or (H2) of Theorem 1.1 of Consonni and Veronese. According to this and to the linearity of the conditional expectation of the mean parameter of $F(\mu)$, we deduce that this NEF is a quadratic. It follows that $F(\nu) = T_\beta(F(\mu))$ is a cubic NEF.

□

In the following theorem, we give a second characterization of the Letac-Mora class of real cubic NEFs.

Theorem 2.6 *Let ν be in $\mathcal{M}(I\!R)$. $F(\nu)$ is cubic if and only if there exist β in $B_{F(\nu)}$ and $(a, b, c) \in I\!R^3$ such that for all m in $M_{F(\nu)}$,*

$$V_{F(\nu)}(m) = (1+\beta m)^3\ \exp\left(a\psi_\nu(m) + bk_\nu(\psi_\nu(m)) + c\right). \tag{2.13}$$

Note that (2.13) may be expressed in terms of the cumulant function as there exist β and $(a, b, c) \in I\!R^3$ such that for all θ in Θ_β

$$k''_\nu(\theta) = (1+\beta k'_\nu(\theta))^3\ \exp\left(a\theta + bk_\nu(\theta) + c\right),$$

that is the cumulant function is solution of some Monge-Ampère equation (see[12]). **Proof** Suppose that $F(\nu)$ is cubic, then there exist β in $B_{F(\nu)}$ and a quadratic NEF $F(\mu)$ such that $F(\nu) = T_\beta(F(\mu))$ or equivalently $F(\mu) = T_{-\beta}(F(\nu))$. Then it follows from (2.9) that

$$V_{F(\nu)}(m) = (1+\beta m)^3 \; V_{F(\mu)}\left(\frac{m}{1+\beta m}\right).$$

It is known (see [1]) that for the quadratic NEF $F(\mu)$ there exists $(a', b', c') \in \mathbb{R}^3$ such that for all m' in $M_{F(\mu)}$

$$V_{F(\mu)}(m') = \exp(a'\psi_\mu(m') + b' k_\mu(\psi_\mu(m')) + c').$$

Writing (2.11) in terms of the mean parameters we get

$$\left\{\begin{array}{ccc} k_\mu\left(\psi_\mu(\frac{m}{1+\beta m})\right) & = & k_\nu(\psi_\nu(m)) \\ \\ \psi_\mu\left(\frac{m}{1+\beta m}\right) & = & \beta k_\nu(\psi_\nu(m)) + \psi_\nu(m) \end{array}\right. \tag{2.14}$$

Therefore

$$V_{F(\nu)}(m) = (1+\beta m)^3 \; \exp(a\psi_\nu(m) + bk_\nu(\psi_\nu(m)) + c),$$

with $a = a', \quad b = b' + \beta a' \quad and \quad c = c'$.
Conversely, if (2.13) holds, then

$$\ln V_{F(\nu)}(m) = 3\ln(1+\beta m) + a\psi_\nu(m) + bk_\nu(\psi_\nu(m)) + c.$$

Taking the derivative, we deduce that the variance function satisfies the differential equation

$$(1+\beta m)V'_{F(\nu)}(m) - 3\beta V_{F(\nu)}(m) = (a+bm)(1+\beta m).$$

Solving this equation by standard methods gives

$$V_{F(\nu)}(m) \quad = \quad \lambda(1+\beta m)^3 + \frac{b}{\beta^2}(1+\beta m)^2 - \frac{b-\beta a}{2\beta^2}(1+\beta m),$$

which is a polynomial of degree less than or equal to 3. □

A third characterization of the cubic NEF's is based on a relation between the associated families of prior distributions Π^β and $\widetilde{\Pi}^\beta$.

Theorem 2.7 *Let ν be in $\mathcal{M}(\mathbb{R})$. $F(\nu)$ is cubic if and only if there exist β in $B_{F(\nu)}$ such that*

$$k'_\nu(\Pi^\beta) \subset \widetilde{\Pi}^\beta.$$

Proof Suppose that $F(\nu)$ is cubic, then from Theorem 2.6 there exist β in $B_{F(\nu)}$ and (a, b, c) in $\mathbb{R}^3$ such that

$$V_{F(\nu)}(m) = (1+\beta m)^3 \; \exp\left(a\psi_\nu(m) + bk_\nu(\psi_\nu(m)) + c\right). \tag{2.15}$$

Consider the set

$$\Omega = \{(t_1, m_1) \in \mathbb{R}_+^* \times \; (M_{F(\nu)})_\beta \; ; t_1 - b > 0 \; and \; \frac{t_1 m_1 + a}{t_1 - b} \in \; (M_{F(\nu)})_\beta\}.$$

We will show that $k'_\nu(\Pi^\beta) = \{\widetilde{\pi}^\beta_{t_1,m_1};\ (t_1,m_1) \in \Omega\}$, which is a part of $\widetilde{\Pi}^\beta$.
Let $t > 0$ and m_0 in $(M_{F(\nu)})_\beta$, and denote by σ the image by k'_ν of the prior π^β_{t,m_0} on θ defined in (2.12). We easily verify that

$$\sigma(dm) = C^\beta_{t,m_0}\ (1+\beta m)\ V^{-1}_{F(\nu)}(m)\ \exp(tm_0\psi_\nu(m) - tk_\nu(\psi_\nu(m)))\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)dm.$$

This using (2.15) becomes

$$\sigma(dm) = \widetilde{C}^\beta_{t_1,m_1}\ (1+\beta m)^{-2}\exp(t_1m_1\psi_\nu(m) - t_1k_\mu(\psi_\nu(m)))\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)dm,$$

with $t_1 = t+b$, $m_1 = \dfrac{tm_0 - a}{t+b}$ and $\widetilde{C}^\beta_{t_1,m_1} = C^\beta_{t_1-b,(t_1m_1+a)/(t_1-b)}$.
We have that $t_1 - b = t > 0$ and $\dfrac{t_1m_1+a}{t_1-b} = m_0 \in (M_{F(\nu)})_\beta$, that is $(t_1,m_1) \in \Omega$. Hence

$$k'_\nu(\Pi^\beta) \subset \{\widetilde{\pi}^\beta_{t_1,m_1};\ (t_1,m_1) \in \Omega\}.$$

In the same, we verify that

$$\{\widetilde{\pi}^\beta_{t_1,m_1};\ (t_1,m_1) \in \Omega\} \subset k'_\nu(\Pi^\beta).$$

Finally, we obtain that

$$k'_\nu(\Pi^\beta) = \{\widetilde{\pi}^\beta_{t_1,m_1};\ (t_1,m_1) \in \Omega\} \subset \widetilde{\Pi}^\beta$$

Conversely, suppose that $k'_\nu(\Pi^\beta) \subset \widetilde{\Pi}^\beta$. The image of an element π^β_{t,m_0} of Π^β by k'_ν is by the very definition

$$k'_\nu(\pi^\beta_{t,m_0})(dm) = C^\beta_{t,m_0}\ (1+\beta m)\ (V_{F(\nu)}(m))^{-1}\ \exp(tm_0\psi_\nu(m) - tk_\nu(\psi_\nu(m)))\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)dm.$$

Since it is assumed to be in $\widetilde{\Pi}^\beta$, there exists (t_1,m_1) in $\mathbb{R}^*_+ \times (M_{F(\nu)})_\beta$ such that

$$k'_\nu(\pi^\beta_{t,m_0})(dm) = \widetilde{C}^\beta_{t_1,m_1}\ (1+\beta m)^{-2}\ \exp(t_1m_1\psi_\nu(m) - t_1k_\nu(\psi_\nu(m)))\ \mathbf{1}_{(M_{\mathbf{F}})_\beta}(m)dm.$$

Comparing these two expressions of $k'_\nu(\pi^\beta_{t,m_0})$ gives

$$V_{F(\nu)}(m) = (1+\beta m)^3\ \exp(a\psi_\nu(m) + bk_\nu(\psi_\nu(m)) + c),$$

where

$$a = tm_0 - t_1m_1,\ b = t_1 - t,\ \textit{and}\ c = \ln\left(\frac{C^\beta_{t,m_0}}{\widetilde{C}^\beta_{t_1,m_1}}\right),$$

According to Theorem 2.6, this is the desired result and the proof is complete. $\square$

3 Example

In this section we illustrate our results by an example involving the most famous family with variance function of degree 3 which is the inverse Gaussian natural exponential family. Consider the distribution

$$\nu(dx) = \frac{(1+x)^{-3/2}}{\sqrt{2\pi}} \; \exp(-\frac{x^2}{2(1+x)}) \; \mathbf{1}_{]-\mathbf{1},+\infty[}(x)\,dx,$$

which is up to an affine transformation an inverse Gaussian distribution. The NEF generated by ν is given by

$$F(\nu) = \{\exp((1+x)\theta + \sqrt{-2\theta}) \; \nu(dx); \; \theta < 0\}, \tag{3.16}$$

its mean parametrization is

$$F(\nu) = \{\exp(-\frac{1+x}{2(1+m)^2} + \frac{1}{1+m}) \; \nu(dx); \; m > -1\}. \tag{3.17}$$

For all $m > -1$ the variance function is given by

$$V_{F(\nu)}(m) = (1+m)^3. \tag{3.18}$$

Let $\mu = T_{-\beta}(\nu)$, we have

$$\mu(dx) = \frac{1}{\sqrt{2\pi}\sqrt{1-\beta x + x}} \; \exp(-\frac{x^2}{2(1+(1-\beta)x)}) \; \mathbf{1}_{\{\mathbf{1}+(\mathbf{1}-\beta)\mathbf{x})>\mathbf{0}\}}(x)\,dx$$

Now for $\beta \neq 0$ we have

$$P(\beta,\theta,\nu)(dx) = \frac{e^{\theta+\sqrt{-2\theta}}}{\sqrt{2\pi}\sqrt{1-\beta x + x}} \; \exp((\theta+\beta(-\theta-\sqrt{-2\theta}))x - \frac{x^2}{2(1-\beta x + x)}) \; \mathbf{1}_{\{1-\beta x+x>0\}}(x)\,dx,$$

so that

$$F^\beta = T_{-\beta}(F(\nu)) = \{P(\beta,\theta,\nu)(dx); \; \theta < 0\},$$

The corresponding family Π^β of conjugate prior distributions is the family of distributions

$$\pi^\beta_{t,m_0}(d\theta) = C^\beta_{t,m_0} \; (1+\beta(\frac{1}{\sqrt{-2\theta}} - 1)) \; \exp\{tm_0\theta + t(\theta + \sqrt{-2\theta})\} \; \mathbf{1}_{(\boldsymbol{\Theta})_\beta}(\theta)\,d\theta,$$

where

$$\begin{aligned}
(\Theta)_\beta &= \;]-\frac{1}{2}(\frac{\beta}{\beta-1})^2 \; , 0[, \\
(M_{F(\nu)})_\beta &= \left\{ \begin{array}{ccc}]-1/\beta, +\infty[& if & \beta < 0 \\ ]-1, +\infty[& if & \beta = 0 \\ ]inf(-1, -1/\beta), +\infty[& if & \beta > 0 \end{array} \right.
\end{aligned}$$

and

$$(C^\beta_{t,m_0})^{-1} = -\frac{\beta}{t}\,[1-\exp(-\frac{tm_0}{2}(\frac{\beta}{\beta-1})^2 + \frac{t\beta}{\beta-1} - \frac{t}{2}(\frac{\beta}{\beta-1})^2)] - (1+\beta m_0)\;\frac{1}{\frac{t}{\beta} - \frac{tm_0}{2}(\frac{\beta}{\beta-1})^2}.$$

defined for $t > 0$ and m_0 in $(M_{F(\nu)})_\beta$
Also, in this example, the family $\widetilde{\Pi}^\beta$ is the set of distributions defined for $t_1 > 0$ and m_1 in $(M_{F(\nu)})_\beta$ by

$$\widetilde{\pi}^\beta_{t_1,m_1}(dm) = \widetilde{C}^\beta_{t_1,m_1}\ (1+\beta m)^{-2}\ \exp(-\frac{t_1(m_1+1)}{2(1+m)^2} + \frac{t_1}{1+m})\ \mathbf{1}_{(M_{\mathbf{F}(\nu)})_\beta}(m)dm,$$

To see how Theorem 2.6 holds in this example, we need only to take $\beta = 1$. Then $\mu = T_{-1}(\nu)$ is the standard gaussian distribution with $V_{F(\mu)}(m') = 1$ for $m' \in \mathbb{R}$. We see that

$$V_{F(\nu)}(m) = (1+m)^3\ \exp(a\psi_\nu(m) + bk_\nu(\psi_\nu(m)) + c),$$

with $a = b = c = 0$. In fact

$$V_{F(\mu)}(m') = \exp(a'\psi_\mu(m') + b'k_\mu(\psi_\mu(m')) + c'),$$

with $a' = b' = c' = 0$, and using the relations $a = a'$, $b = b' + \beta a'$, $c = c'$, we get $a = b = c = 0$.
Concerning Theorem 2.7, we fist observe that the hypotheses in this theorem are well verified. In fact, let π^1_{t,m_0} be the prior on the natural parameter θ and $\widetilde{\pi}^1_{t_1,m_1}$ be the prior on the mean parameter m, then

$$\begin{aligned} t_1 &= t + b = t > 0 \\ m_1 &= \frac{tm_0 - a}{t+b} = m_0 \in\ (M_{F(\nu)})_1 =]1, +\infty[. \end{aligned}$$

The density function of π^1_{t,m_0} is equal to

$$\frac{1}{\sqrt{-2\theta}}\ \exp\{tm_0\theta + t(\theta + \sqrt{-2\theta})\},$$

and for all $m > -1$ the density function of $k'_\nu(\pi^1_{t,m_0})$ is given by

$$(1+m)^{-2}\ \exp\{-\frac{tm_0}{2(1+m)^2} - t(\frac{1}{2(1+m)^2} - \frac{1}{1+m})\}$$

which is equal to $\widetilde{\pi}^1_{t,m_0}$.

References

[1] Casalis, M. (1996). The 2d+4 simple natrual exponential families on $\mathbb{R}^d$, *Ann. Statist.* **24** 1828-1854.

[2] Cifarelli, D.M. and Regazzini, E. (1983). Qualche Osservazione sull'Uso di distribuzioni iniziali coniugate alla famiglia esponenziale, *Statistica*, **43**, 415-424. Corrections: *Statistica*(1990), 50, 293.

[3] Consonni, G. and Veronese, P. (1992). Conjugate priors for exponential families having quadratic variance functions,
Journal of the American Statistical Association, **87**, 1123-1127.

[4] Diconis, P and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist* **7** 269-281.

[5] Feinsilver, P.(1986). Some classes of orthogonal polynomials associated with martingales, *Proc. Amer. Math. Soc.* **98** 298-302.

[6] Hassairi, A. (1992). La classification des familles exponentielles naturelles sur $\mathbb{R}^d$ par l'action du groupe linaire de $\mathbb{R}^{n+1}$, *C.R.Acad.Sc.* **315**.

[7] Hassairi, A. Kokonendji, C., and Masmoudi, A. (2004). Implicit distribution and estimation, *Communication in Statistics*, **34** 2, 245-252.

[8] Hassairi, A. and Zarai, M. (2004). Characterization of the cubic exponential families by orthogonality of polynomials, *Annals of probability*, **32** 2463-2476.

[9] Letac, G. Lectures on natural exponential families and their variance function. Monografias de Matemática. 50, IMPA, Rio de Janeiro.

[10] Letac, G. and Mora, M. (1990). Natural real exponential familiies with cubic variance functions, *Ann. Statist*, **18** 1-37.

[11] Morris, C. N. (1982). Natrual exponential families with qudratic variance functions. *Ann. Statist.* **10** 65-80.

[12] Zuily, C. (1989). Local existence and regularity of the Dirichlet problem for the Monge-Ampère equation, Journes "équations aux Drives Partielles" (Saint Jean de Monts, 1989), Exp. No. XXI, 5 pp.